ORIGINAL PAPER

cDNA-AFLP-based genetical genomics in cotton fibers

**Michel Claverie · Marlène Souquet · Janine Jean · Nelly Forestier-Chiron ·
Vincent Lepitre · Martial Pré · John Jacobs · Danny Llewellyn · Jean-Marc Lacape**

Abstract Genetical genomics, or genetic analysis applied
to gene expression data, has not been widely used in plants.
We used quantitative cDNA-AFLP to monitor the variation
in the expression level of cotton fiber transcripts among
a population of inter-specific *Gossypium hirsutum* ×
G. barbadense recombinant inbred lines (RILs). Two key
fiber developmental stages, elongation (10 days post
anthesis, dpa), and secondary cell wall thickening (22 dpa),
were studied. Normalized intensity ratios of 3,263 and
1,201 transcript-derived fragments (TDFs) segregating
over 88 RILs were analyzed for quantitative trait loci
(QTL) mapping for the 10 and 22 dpa fibers, respectively.
Two-thirds of all TDFs mapped between 1 and 6 eQTLs
(LOD > 3.5). Chromosome 21 had a higher density of
eQTLs than other chromosomes in both data sets and,
within chromosomes, hotspots of presumably *trans*-acting

eQTLs were identified. The eQTL hotspots were compared
to the location of phenotypic QTLs for fiber characteristics
among the RILs, and several cases of co-localization were
detected. Quantitative RT-PCR for 15 sequenced TDFs
showed that 3 TDFs had at least one eQTL at a similar
location to those identified by cDNA-AFLP, while 3 other
TDFs mapped an eQTL at a similar location but with
opposite additive effect. In conclusion, cDNA-AFLP
proved to be a cost-effective and highly transferable plat-
form for genome-wide and population-wide gene expres-
sion profiling. Because TDFs are anonymous, further
validation and interpretation (in silico analysis, qPCR gene
profiling) of the eQTL and eQTL hotspots will be facili-
tated by the increasing availability of cDNA and genomic
sequence resources in cotton.

Communicated by G. Bryan.

M. Claverie · M. Souquet · V. Lepitre · M. Pré ·
J.-M. Lacape (✉)
UMR AGAP, CIRAD, Avenue Agropolis,
34398 Montpellier, France
e-mail: marc.lacape@cirad.fr

J. Jean · N. Forestier-Chiron
UR SCA, CIRAD, Avenue Agropolis,
34398 Montpellier, France

J. Jacobs
Bayer BioScience N.V., Technologiepark 38, Ghent, Belgium

D. Llewellyn
CSIRO Plant Industry, P.O. Box 1600, Canberra, ACT, Australia

Introduction

A longstanding goal in plant genetics is to understand the
molecular cause of the variation in complex traits. Quan-
titative trait loci (QTL) mapping, as the most common
approach, has in past years described numerous chromo-
somal regions correlated with a wide array of traits, rang-
ing from biotic/abiotic stress tolerance to yield and quality
components. However, in spite of these considerable
efforts, the underlying genes responsible for particular
quantitative traits remain largely unknown and only in a
few examples has it been possible to identify and clone the
genes responsible for conferring a trait (Salvi and Tuberosa
2005).

The ability to study gene expression variation at the
individual, population, and species level has recently been
enhanced by genomic methodologies such as transcripto-
mics. Jansen and Nap (2001) introduced the concept of

"genetical genomics" whereby different medium and high-throughput transcriptomic technologies enable the mapping of expression QTL (eQTL), e.g. of the variability in the expression levels of a gene(s) over a segregating mapping population. The gene expression level, quantified by the abundance of its mRNA, is considered as a 'phenotype' (sometimes referred as an e-trait), possibly influenced by genetic determinants. Classical genetic analysis (comparison of the mean phenotypic values of the e-trait between the genotypic classes of the segregating population) is used to map QTLs and can therefore allow the identification of the genes and/or regulatory regions that control expression phenotypes.

Beyond single gene expression measurement (northern blotting, RT-qPCR), different methods of profiling large numbers of genes have emerged, including cDNA or oligonucleotide microarrays, tag-based methods (SAGE, MPSS) and PCR-based differential profiling techniques, including differential display and cDNA-AFLP. The latter appears as a valuable and cheap alternative for functional genomics and genetics, as compared to hybridization-based microarray techniques. cDNA-AFLP was also shown to increase the resolution of expression pattern detection using smaller amounts of mRNA (Reijans et al. 2003). In particular, the expression patterns visualized by cDNA-AFLP have been shown to correlate well with northern blot analyses (Albertini et al. 2004; Bachem et al. 1996; Donson et al. 2002).The quantitative response in the cDNA-AFLP system therefore seems to be sufficiently precise and broadly proportional to the input DNA (Bachem et al. 1996; Breyne et al. 2003; Vuylsteke et al. 2006) for demanding applications like transcriptional profiling and eQTL mapping. cDNA-AFLP has been used in cotton previously to compare the transcriptomes of two cotton lines (one fertile and the other male sterile) (Ma et al. 2008), to identify genes involved in somatic embryogenesis (Leng et al. 2007) and to study gene silencing (Adams et al. 2004). In a recent report (Liu et al. 2011), cDNA-AFLP was used for mapping purposes in cotton.

Global analysis of gene expression followed by eQTL mapping has been reported in human and animal systems (Gibson and Weir 2005; Gilad et al. 2008; Hubner et al. 2006), in yeast (Brem et al. 2002), and to a lesser extent in plants (Holloway and Li 2010; Kliebenstein 2008). Microarray profiling followed by eQTL mapping in plants has been reported in model crops like Arabidopsis (Keurenjes et al. 2007; West et al. 2007) and rice (Wang et al. 2010). Four other eQTL reports from non-model crops, include barley (Potokina et al. 2008), maize (Shi et al. 2007), wheat (Jordan et al. 2007), and eucalyptus (Kirst et al. 2005). In only one case has cDNA-AFLP been used to map expression polymorphisms over a population of 50 Arabidopsis recombinant inbred lines (RILs) from the cross Ler × Cvi (Vuylsteke et al. 2006).

Differences in expression of a given gene may result either from allelic differences in its promoter, non-coding, or coding regions or from effects of distal regulatory loci. In other words, an eQTL may map to the genetic position of the gene itself, indicating that *cis* changes are responsible for the different levels of expression. In contrast, genes revealing eQTLs at positions different from the gene are thought to be regulated by *trans*-acting factors. Evidence of both *cis* expression polymorphism as well as clustered *trans*-eQTL have been reported in a number of organisms including yeast and Arabidopsis (Gibson and Weir 2005; Keurenjes et al. 2007; West et al. 2007). Several authors have questioned the relevance of the commonly reported phenomenon of eQTL co-localizing in hotspots (Breitling et al. 2008; Michaelson et al. 2009; Pérez-Enciso et al. 2007) recommending caution in interpretation because of inter-correlations between transcript levels which can be related to interconnected pathways or to fortuitous associations (interactions with environment or with technical artifacts, or population structure).

There are two economically important tetraploid species of cultivated cotton, *Gossypium hirsutum* ("Upland" cotton) and *G. barbadense* (Caribbean "Sea-Island", Extra Long Staple "ELS", "Pima" and "Egyptian" cottons). They display many complementary agronomic features and are widely interbred in cotton breeding programs. *G. hirsutum* (hereafter *Gh*), the most widely cultivated species, has higher yield potential than *G. barbadense* (*Gb*) in most environments; however, *Gb* cultivars are superior to *Gh* in most aspects of fiber quality, such as fiber length, strength and fineness. Molecular data indicate that the two species share a common allopolyploidization origin (∼1–2 MYA) between an A-genome and a D-genome diploid species (Wendel and Cronn 2003).

Cotton fibers are highly elongated single cells of the epidermal layer of the ovule. Fiber development spans four discrete, yet overlapping stages: initiation [−3 to 5 days post anthesis (dpa)], elongation (3–21 dpa), secondary cell wall (SCW) deposition (14–45 dpa) and maturation/dehydration (40–55 dpa) (Basra and Malik 1984). Their commercial value is determined by their overall physical dimensions and the extent of thickening of the internal walls, properties that affect yarn-spinning and other fabric manufacturing processes. Although earlier genetic studies have demonstrated fairly high heritabilities for fiber characteristics (May 1999), the various fiber quality QTL mapping efforts reported so far indicate a very complex regulation of cotton fiber quality as a whole (Lacape et al. 2010; Rong et al. 2007). We recently reported the QTL analysis across 11 environments (meta-analysis) of the same interspecific *Gh* × *Gb* RIL population studied here and its comparison with other fiber QTL reports (Lacape et al. 2010). Although congruence between environments

was only partial, evidence was provided for a co-localization of phenotypic QTLs (phQTL) in some regions of the genome.

Although considerable functional genomics work on cotton fiber development has been carried out, it has essentially been conducted within the *G. hirsutum* species, including its derived fiber mutants. Very little information has been accumulated on the differences in fiber gene expression between the two species *Gh* and *Gb*. It is only recently that two studies specifically focused on the differences between *Gh* and *Gb* at the transcriptional level using microarrays (Al-Ghazi et al. 2009; Alabady et al. 2008). Al-Ghazi et al. (2009) reported that, under the conditions in which they were grown and despite their final fiber physical differences, the two species were fairly synchronous in fiber elongation and cellulose accumulation during SCW thickening. This last point implies that comparing *Gh* and *Gb* transcripts at a given time point of fiber development accounts for differences in gene regulation rather than differences in fiber phenology.

This report extends our previous studies in which an interspecific *G. hirsutum* × *G. barbadense* RIL population has served for genetic mapping (Lacape et al. 2009) and for fiber QTL meta-analysis (Lacape et al. 2010). We have conducted a fiber transcriptome study from a geneticist's perspective: we analyzed genome-wide gene expression variation using cDNA-AFLP on the same RIL population. The specific objectives were to (1) assess the applicability of cDNA-AFLP as a technique for population-wide expression profiling and eQTL mapping of transcripts in developing fibers, and (2) to conduct comparative mapping approaches to test the coincidence of expression-derived QTLs (eQTLs) and phenotype-derived QTLs (phQTLs).

Materials and methods

Plant material

mRNA from developing fibers of an interspecific *G. hirsutum* × *G. barbadense* RIL population was used for gene expression analysis. A hundred RILs in the F_6–F_8 generation of single seed descent and the two parents, Guazuncho 2 (*G. hirsutum*) and VH8-4602 (*G. barbadense*) were grown in a glasshouse at Montpellier (France) from December 2007 through to August 2008. All genotypes (RILs and parents) were grown in a 10 l pot as two plants per pot (2 replicate pots for the 2 parents) and the 104 pots were randomly distributed. Growth conditions in the glasshouse were non-limiting in terms of nutrition and water availability (twice daily drip irrigation), while light was adjusted to simulate normal (tropical) day/night cycles and a reversible heating/cooling system minimized

temperature variation during the sampling period (variation 25–35°C day/night).

Flowers were tagged on the day of anthesis. Two fiber developmental time points were chosen: 10 dpa, corresponding to the phase of peak elongation of fibers and 22 dpa corresponding to the transition phase from primary to SCW synthesis. Developing bolls (each containing 3–5 locules), 5–10 bolls per plant for 10 dpa and 2–5 bolls per plant for 22 dpa, were harvested between 10.00 am and 2.00 pm from the three different plants per RIL and parent. Boll coats were removed and whole locules immediately plunged into liquid nitrogen before storage at −80°C. Ovules from both parental lines were also collected to be used as control RNAs.

RNA extraction

Fiber total RNA was extracted for each RIL and parent from a pooled sample of several bolls from the two plants: a total of 3–4 (10 dpa fibers) or 2–3 (22 dpa fibers) locules each from a different boll were randomly selected from both plants and pooled before grinding. Seeds were separated from fibers during the first step of grinding in liquid nitrogen. Fibers (1–2 g) were ground mixed with 0.5–1 g polyvinylpolypyrrolidone and the powder kept frozen at −80°C for no more than one week before processing for RNA extraction. For the 10 dpa fibers, total RNA was extracted following the protocol described in Argout et al. (2008). This protocol, based on MATAB extraction and LiCl precipitation is rapid and suitable for small scale fiber RNA extraction but yielded less RNA and was more heterogeneous for the 22 dpa. RNA from the 22 dpa fibers was therefore extracted according to the more labor intensive protocol from Wan and Wilkins (1994) using the Hot Borate RNA extraction buffer E.

RNA quality was checked on 1.4% denaturing agarose gel. Total nucleic acids were quantified by UV absorbance (DU530, Beckman Coulter, Fullerton, CA, USA) and DNA contaminations were quantified by fluorescence of the dye Hoechst 33258 with a Fluoroskan fluorimeter (Ascent, Labsystems, Finland).

cDNA-AFLP protocol

The cDNA-AFLP protocol used (detailed in Online Resource ESM1) was a modification of the original procedure of Bachem et al. (1996). This technique as further detailed in Vuylsteke et al. (2007) is characterized by (i) the generation of a single cDNA fragment for each messenger ('one-gene-one-tag') originally present in the sample (Breyne et al. 2003), (ii) the combination of a five-cutter and a four-cutter restriction enzymes, (iii) the production of cDNA-AFLP bands that are derived from the 3′-end of the gene. Three enzyme combinations were

originally tested: *Apo*I/*Mse*I, *Taq*I/*Mse*I and *BstY*I/*Mse*I. The *BstY*I/*Mse*I combination was chosen because it produced an optimal density of bands after polyacrylamide gel electrophoresis (using *BstY*I-C + 1 and *Mse*I + 2 primers). The protocol of Vuylsteke et al. (2007) was modified as follows: cDNA synthesis using the bead anchored oligo (dT) as primer, reduction of the number of PCR cycles during pre-amplification and selective amplification runs, and higher amounts of DNA template (ligation or PCR product) used in the PCR. After pre-amplification, the mixture was used for selective amplification with the 62 and 64 selective primer combinations for the 10 and 22 dpa experiments, respectively.

In order to simultaneously electrophorese all of the studied RILs and parents on the same gel (94 slots) and to allow for controls and the marker ladder, only a subset of 88 RILs could be analyzed and these were chosen on the basis of their genomic content (RILs that were heavily biased in their allelic content were discarded) as were those for which the quantity or quality of the RNA were low. The migration of the cDNA-AFLP bands, later referred to as TDFs were compared to a 30–330 bp size reference marker (Invitrogen). The sizes of the fragments larger than 330 bp were therefore only approximate.

Analysis of digital AFLP gel images and generation of normalized expression data

Autoradiograms (obtained from 62 and 64 *BstY*I-CN/*Mse*I-NN combinations for the 10 and 22 dpa samples, respectively) were scanned at 300 DPI and 16 bits gray format. All images were then quantified for band intensity using the QuantarPro software (Keygene N.V., Wageningen, The Netherlands). For each gel, vertical alignment bands were automatically defined and then manually adjusted to match the lanes. As the fiber was removed from whole ovules during the initial grinding phase there was potential for the fiber material to be contaminated with small amounts of ovule tissue. This was checked using parental ovule-derived cDNA-AFLP profiles as comparisons: the few bands more intense in the ovule control than in the fiber samples were then not scored. All the bands were automatically quantified and any misalignments corrected after visual inspection. Depending on the quality of the gel, between 50 and 80% of the bands were reliably scored. Qualitative polymorphisms (segregating as presence–absence of AFLP bands), probably associated to SNPs/indels, were observed in a limited extent (<3%, not shown). These bands were not considered for quantification.

The AFLP radiograms of the two series of samples, 10 and 22 dpa, were tentatively aligned to find any transcripts common between the experiments, but this was only possible in around 10% of the primer pairs (not shown).

The non-normalized quantitative data were then imported into an Excel spreadsheet. The mean-normalization of the band intensity signals was carried out in two steps. First, for each TDF, intensity signals for each RIL were divided by the mean intensity of the TDF over the whole RIL population ("horizontal" normalization). Then, for a given RIL, the resulting normalized intensity of each individual TDF was divided by the mean intensity value calculated from the raw intensity signals of all the TDFs present on the gel (or gel part) in each individual RIL ("vertical" normalization). Due to differences between the top and the bottom of the gels (i.e. number and thickness of bands, background signal), the 10 dpa gels were split in two parts (above and below 250 bp) and the "vertical" normalization was carried out separately.

Characterization of AFLP fragments: isolation and cloning of TDFs

Validation of the cDNA-AFLP-based eQTL mapping was undertaken for some cloned TDFs by quantitative RT-PCR. TDFs were re-amplified from 10 dpa cDNA samples of three separate RILs (chosen from among the high-expressing RILs), and using the same protocol described above. Five cDNA-AFLP gels were re-run using the five different AFLP primer pairs appropriate for the particular TDFs. A total of 29 bands (see "Results") were cut from the gels and DNA eluted overnight in 100 μl of sterile water. The TDFs were re-amplified using the same PCR primers and conditions and ligated in the pCR2-TOPO vector (Invitrogen), and transformed into *Escherichia coli*. For each TDF, three individual clones were isolated and sequenced. The nucleotide sequences were compared with publicly available cotton EST databases by BLAST sequence alignments (Altschul et al. 1990). Lists of cloned TDFs, primers and other features are summarized in Online Resource ESM2.

Quantitative RT-PCR analysis

The cDNAs of the same 88 RILs (10 dpa fibers) used in the cDNA-AFLP profiling experiment were synthesized from the same RNA samples. For each RIL, 2 μg of total RNA were dried in a vacuum and resuspended in 12 μL of water and 2.5 μL oligodT(23)VN 100 μM. The mixture was denatured for 10 min at 65°C. Then, first strand cDNA synthesis was performed in 1× Expand Reverse Transcriptase buffer, 10 mM DTT, 1 mM dNTP (each), 20 U of RNAse inhibitor and 50 U of Expand Reverse Transcriptase (Roche). Reactions were incubated at 47°C for 1 h, and then the cDNAs were diluted tenfold in water to use as templates for quantitative RT-PCR (qPCR). qPCR was performed in a LightCycler 480 (Roche). Each qPCR

reaction was carried out in triplicate in a total volume of 15 μL that contained 1x SYBR Green I Master Mix (Roche) and 3 μL of tenfold diluted cDNA template per reaction, plus 500 nM of each forward and reverse gene-specific primer (Online Resource ESM2). Cycle conditions were: 95°C for 5 min and then 95°C for 20 s, 60°C for 15 s, 72°C for 20 s (45 cycles). A melting-curve analysis was carried out to evaluate the specificity of the primers as recommended (Roche). Primers for qPCR were designed using Primer3 and default parameters. Whenever possible, primers were designed to amplify a fragment in the 3′UTR region of the target gene. For some TDFs, for which the sequence Blast result could not discriminate between several putative ESTs, primer pairs were designed for each EST. The amplification efficiencies of all the primers pairs were tested on serial dilutions of cDNAs from each of the two parents. Primer efficiencies were near 100% for both genotypes. Quantification of gene expression was carried out using the $\Delta\Delta C_t$ method of relative quantification, based on the differences of C_t values between the reference gene Ubiquitin ligase and the target gene tested.

Linking cDNA-AFLP and EST sequence data

EST resources from GeneBank (assembly cotton46 from http://cottonevolution.info/ as a hybrid assembly combining both Sanger and NGS EST sequences) and from our own fiber cDNA pyrosequencing project (ESTs from this study, unpublished data) were assembled together in a new unigene set using MIRA software (unpublished). The resulting unigene set, consisting of 38,297 contigs (singletons excluded), was searched in silico for the presence of *BstYI/Mse*I AFLP fragments using the *cDNA-AFLP* module of the AFLP-in silico program (Rombauts et al. 2003). For each input sequence, only a single fragment (noted pTDF for predicted TDF) is generated corresponding to the fragment in the most 3′-terminal position flanked by the two restriction sites in 5′ (RGATCC for enzyme *BstY*I) and 3′ (TTAA for enzyme *Mse*I) end positions, respectively; in accordance with the principle of 'one-gene-one-tag' underlying the cDNA-AFLP method that we followed here (Breyne et al. 2003; Vuylsteke et al. 2007). An Excel VBA program was then used to search for matches between the TDFs quantified on the cDNA-AFLP gels using QuantarPro for eQTL mapping and the pTDFs predicted in silico. The search for matches was based upon (i) the same combination of flanking nucleotides (64 possible combinations of *BstY*I-CN and *Mse*I-NN selective nucleotides) and (ii) the similar product size in number of nucleotides. Because band resolution is not uniformly distributed on a polyacrymalide gel, we allowed for an error margin depending on the position of the fragment on the gel: ±1 nt in range 50–149 bp, ±2 nt in range

150–249 bp, and ±3 nt in range 250–340 bp (Qin et al. 2006).

The ESTs developed in our project (unpublished) were partly derived from the 454 pyrosequencing of four non-normalized fiber cDNA libraries from the two parents of the RIL population (Guazuncho 2 and VH8) and for the two fiber development dates under study for eQTL mapping (10 and 22 dpa). The normalized number of reads of the respective four libraries in each unigene (depth) was then used to compute digital expression levels (Torres et al. 2010). An absolute fold change greater than 2, associated with a *q* value lower than 0.05 (Guo et al. 2010) between two conditions, either Gh10 versus Gb10 or Gh22 versus Gb22, was then used to identify digitally differentially expressed genes and pTDFs; for comparison with cDNA-AFLP-derived TDFs that mapped eQTLs.

QTL analysis

WinQTL Cartographer software (Basten et al. 2003) was used to perform simple marker analysis (SMA), interval (IM) and composite interval mapping (CIM) on the 4,464 TDFs derived from the cDNA-AFLP profiles quantitatively scored by QuantarPro and on the genes quantified by qPCR. These analyses were performed over 88 RILs using the genotype data from 656 loci chosen from the original 800 marker data set along the RIL map (2,044 cM) previously reported (Lacape et al. 2009), corresponding to an average density of one marker per 3.1 cM. CIM analysis was applied using Model 6, Forward Regression Method, 5 control markers chosen as cofactors by stepwise regression, a 2 cM walking speed and a window width of 10 cM. The five control markers were selected by WinQTL Cartographer from a first step univariate analysis. A genome-wide empirical permutation test (1,000 permutations) was performed on 2 sets (10 dpa) and 1 set (22 dpa) of 100 randomly chosen TDFs each. Average LOD threshold (global risk of 5%) values were 3.52 ± 0.54 and 3.59 ± 0.88 for the 10 dpa sets and 3.51 ± 0.43 for the 22 dpa set. LOD3.5 was therefore used as a common threshold in all analyses. All peak LOD-positions and their confidence intervals (CI) were exported from WinQTL Cartographer with following parameters: (1) peak LOD positions (greater than 3.5) spaced by at least 5 cM as an exclusion window with a LOD value from top to valley greater than 1, and (2) the interval corresponding to a one-LOD drop-off used as a CI. In addition to the peak LOD positions and their CI, the percent of variation explained (R^2) and the additive effect values were also exported from WinQTL Cartographer. This later parameter was used to test putative transgressive segregation as in West et al. (2007): for those TDFs exhibiting two or more eQTLs, a transcript

was reported as showing transgressive segregation when it mapped at least two eQTLs with parental effects in opposite directions.

eQTL distribution

The support intervals (one-LOD drop off) of all significant eQTLs (LOD > 3.5) were used to calculate the total eQTL representation along bins of fixed length (2 cM) on the RIL map. The 1,020 bins along the 2,044 cM long RIL map were designated by the chromosome name and order on the chromosome. For example, 'bin 2_5' designated the fifth bin, or segment between 8 and 9.99 cM (inclusive), on chromosome 2. A permutation (1,000) test, using 0.05 as a threshold (95th percentile) was performed to determine a threshold value for which chromosome bins were significantly over-populated with eQTLs. These over-populated bin regions were also referred to as eQTL hotspot (Kliebenstein et al. 2006; Potokina et al. 2008).

Comparison of eQTL to phenotypic fiber QTL

The QTL analysis of gene expression data and the QTL mapping of phenotypic fiber data collected from 11 sets of experiments (Lacape et al. 2010) were conducted using the same parameters of WinQTL Cartographer and over the same genotypic data set. However, LOD threshold filtering was applied differently for eQTLs and for phenotypic QTLs (phQTLs). Co-incident localization of phQTLs across experiments was assessed using a low (LOD > 2) threshold and meta-clusters of phQTLs were then identified for a limited number of fiber traits and chromosomes (Lacape et al. 2010). The chromosome regions delineated by phQTL meta-clusters, spanning between 10 and 20 cM, were then used to assess the co-localization with eQTLs using the overlap with their confidence intervals.

Results

TDF detection

cDNA-AFLP was used to analyze global transcript abundance in developing cotton fibers harvested at two developmental stages, 10 and 22 dpa. Each of the selective AFLP primer combinations amplified a range 100–150 fragments.

Image analysis of 126 gel electrophoresis scans allowed the quantification of transcript abundance variation for 3,263 and 1,201 segregating TDFs for the 10 and 22 dpa fibers, respectively. For each AFLP primer pair, between 27 and 89 TDFs were quantitatively scored for the 10 dpa samples (average 54), which represents around 60% of the total number of bands (not shown), and between 7 and 34 for the

22 dpa samples (average 19), which represents around 30% of the total number of bands (not shown). Normalized TDFs signal intensities were then calculated and used in further analysis as described in the "Materials and methods".

eQTL statistics

The SMA method detected 3,962 (for 10 dpa) and 1,404 (for 22 dpa) significant marker × expression significant associations using $Pr(F) < 0.001$ as a threshold. This $Pr(F)$ value roughly corresponded to a LOD threshold of 3.5 in interval mapping (not shown). The CIM method detected 3,665 and 1,375 LOD peaks (LOD > 3.5) or eQTLs. SMA and CIM methods globally agreed in the number and location of significant eQTLs (not shown); only CIM results are presented in detail.

In each of the 10 and 22 dpa data sets, two thirds of all TDFs detected at least one significant eQTL (LOD > 3.5): 68% (2,220 out of 3,263) and 67% (803 out of 1,201) of the 10 and 22 dpa TDFs detected between 1 and 6 eQTLs, respectively (Table 1). As in previously reported whole genome eQTL studies, a majority (85 and 83%) of the transcripts displaying at least one significant eQTL were controlled by just one or two eQTLs (Table 1).

Only a minority had high LOD scores (10% with LOD > 7) (Fig. 1). The R^2 values were distributed in an asymmetrical distribution (Fig. 1). The majority (71 and 79%) of eQTLs had an R^2 between 0.1 and 0.2, while mean R^2 values were of 0.19 and 0.17 for the 10 and 22 dpa experiments, respectively. Distribution of the additive effects as conferred by either *G. hirsutum* or *G. barbadense* parental alleles indicated that 1,670 (45.6%) of all 10 dpa differentially expressed TDFs were up-regulated by parent Guazuncho (Gh) and 1,995 (54.4%) by parent VH8 (Gb). Proportions were similar for the 22 dpa experiment: 679 (49.4%) TDFs were Gh-up-regulated and 696 (50.6%) were Gb-up-regulated.
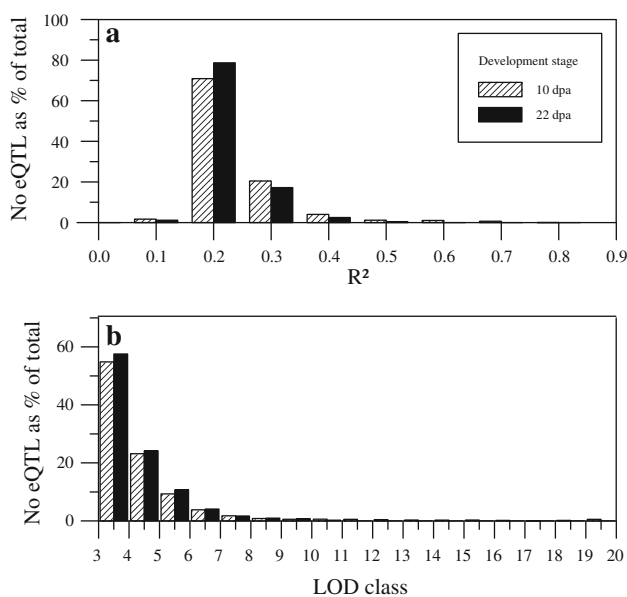
Transgressive segregation was indirectly assessed by the existence, for a given TDF, of eQTLs with opposing parental allelic effects (West et al. 2007). For the 1,025 and 388 TDFs of the 10 and 22 dpa experiments with at least 2 eQTLs, 54 and 59% (553 and 230, respectively) displayed transgression, respectively.

Distribution of eQTLs among chromosomes

Table 2 shows the respective total number of LOD peaks on the 26 chromosomes. The 3,665 and 1,375 eQTLs detected in the 10 and 22 dpa experiments were spread throughout the genome but were not evenly distributed among chromosomes (Table 2). There was a slight although not statistically significant bias for the 10 dpa eQTLs in favor of chromosomes of the A_t sub-genome

Table 1 Number of significant eQTL (LOD > 3.5) detected per TDF across the 10 and 22 dpa experiments

Number of eQTL per TDF	10 dpa		22 dpa	
	Observed number of TDFs	Percentage	Observed number of TDFs	Percentage
0	1,043	32.0	398	33.1
1	1,195	36.6	415	34.6
2	689	21.1	250	20.8
3	266	8.2	103	8.6
4	57	1.7	26	2.2
5	12	0.4	7	0.6
6	1	0.03	2	0.2
Total	3,263		1,201	

Fig. 1 Frequency distribution of eQTLs. Distribution of significant (LOD > 3.5) fiber eQTLs among R^2 (**a**) and LOD classes (**b**) for the 10 and 22 dpa experiments

(c1–c13), 1,925, when compared with the D_t sub-genome (c14–c26), 1,740. The situation was reversed in the case of the 22 dpa eQTLs (627 and 748 on A_t and D_t chromosomes, respectively).

This study clearly highlighted a few chromosomes with a higher representation of eQTLs: c21, with 394 and 129 eQTLs at 10 and 22 dpa stages, respectively, c24 (123 eQTL at 22 dpa), and to a lesser extent c2, c5, c9, c12, and c15 (>190 eQTLs at 10 dpa) (Table 2).

Distribution of eQTLs within chromosomes and eQTL hotspots

In order to obtain a more realistic representation of the distribution of eQTLs within chromosomes, the CI of the

LOD peaks and their overlap over fixed map segments were considered, preferably to their exact peak position. This also buffered for any local map inaccuracies. The total length of the 26 chromosomes of the RIL map, i.e. 2,044 cM (Lacape et al. 2009), was divided into 1,020 bins of 2 cM bins and the position of an eQTL was defined by the bins upon which its one-LOD drop off CI overlapped. The CI of the 3,665 and 1,375 significant eQTLs in the 10 and 22 dpa experiments were on average 6.6 and 5.8 cM, respectively, indicating that an eQTL on average overlapped with four bins of 2 cM; i.e., the 3,665 LOD peak positions of the 10 dpa samples converted into 15,687 hits with a 2 cM bin and the 1,375 LOD peak positions of the 22 dpa samples converted into 5,375 hits. This mode of visualization (Fig. 2) smoothed the variation in eQTL frequency within chromosomes as compared to a strict consideration of the eQTL peak positions (not shown).

eQTLs were clearly non-randomly distributed (Fig. 2). Within a single 2 cM bin, the maximum number of overlapping CI reached 140 for bin 21_36 in the 10 dpa experiment (with 73 LOD peaks effectively mapping within this same bin). Using 33 and 17 eQTLs (overlapping CI) per bin as thresholds ($P < 0.01$), there was an excess of eQTLs over 91 and 41 of the individual bins in the 10 and 22 dpa experiments, respectively. Considering adjacent significant bins, or in 5 cases nearly significant bins, 33 chromosomal segments further referred as eQTL hotspots were identified. Summary details of the 33 hotspots (21 and 12 for the 10 and 22 dpa experiments, respectively) are given in Table 3. The highest number of overlapping eQTL support intervals in a given hotspot was 1,038 eQTLs at 10 dpa mapped on c21 spanning 32 cM, or 16 significant adjacent bins (from bin 21_32 to 21_47). For the 10 dpa experiment, the 21 hotpots contained 1,468 eQTLs (40% of all 3,665 eQTLs), or 4,306 overlapping CI, and covered 91 bins (9% of all 1,020 bins). For the 22 dpa experiment, the 12 hotpots contained 366 eQTLs (27% of all 1,375 eQTLs), or 849 overlapping CI, and covered 41 bins (4% of all 1,020 bins).

Mean LOD and mean R^2 of the eQTLs within hotspots were not different from overall values (Table 3). Within a given hotspot, the direction of additive effects was most often balanced between the two parents (Table 3): only 5 of the 33 were significantly ($P < 0.05$) biased for an over-representation conferred by either Gh (4 cases) or Gb (1) alleles. Three additional cases (all with an over-representation by Gb) were close to significance ($P < 0.1$).

As with the total numbers of significant eQTLs per chromosome, there was also a tendency for the A_t chromosomes (c1–c13) to map more 10 dpa eQTL hotspots than the D_t chromosomes (c14–c26), 13 versus 8, and for the D_t chromosomes to map more 22 dpa eQTL hotspots (9 vs. 3) (Table 3). There was no evidence for homoeologous

Table 2 Number of eQTLs per chromosome in the 10 and 22 dpa experiments

Chromosome		1	2	3	4	5	6	7	8	9	10	11	12	13	Total for A
10 dpa	No. eQTL	119	190	151	181	203	111	126	93	216	65	134	207	129	1,925
	%	3.2	5.2	4.1	4.9	5.5	3.0	3.4	2.5	5.9	1.8	3.7	5.6	3.5	52.5
22 dpa	No. eQTL	44	46	60	32	96	40	38	34	67	27	33	69	41	627
	%	3.2	3.3	4.4	2.3	7.0	2.9	2.8	2.5	4.9	2.0	2.4	5.0	3.0	45.6

Chromosome		14	15	16	17	18	19	20	21	22	23	24	25	26	Total for D	Total
10 dpa	No. eQTL	151	192	113	68	110	175	163	394	57	84	99	76	58	1,740	3,665
	%	4.1	5.2	3.1	1.9	3.0	4.8	4.4	10.8	1.6	2.3	2.7	2.1	1.6	47.5	
22 dpa	No. eQTL	54	21	39	40	56	87	37	129	31	60	123	42	29	748	1,375
	%	3.9	1.5	2.8	2.9	4.1	6.3	2.7	9.4	2.3	4.4	8.9	3.1	2.1	54.4	

A_t/D_t relationship between the localizations of multiple individual eQTLs derived from the same TDF (not shown) or between eQTL hotspots on the different genomes.

Although eQTL hotspots were less numerous in the 22 dpa experiment than in the 10 dpa experiment, the comparison of the distribution of the eQTL in the two data sets showed few instances of conservation in the locations of eQTL hotspots along chromosomes (Figs. 2, 3): bins 17–21 on c2 (slightly shifted) (Fig. 3), bin 11 (possibly also bin 12) of c5, bins 27 and 28 (possibly also bin 29 close to significance) of c9, and bins 75–77 (possibly also bins 74 and 78) of c21 (Fig. 3). However, some of the most significant hotspots were clearly stage-specific: on c2 (bottom), c11, c12, c15, c19 and c21 (central region) for the 10 dpa hotspots, or on c19 and c24 for the 22 dpa eQTL hotspots (Table 3).

Correspondence between eQTLs and phQTLs for fiber quality traits

The fiber quality QTLs from the same RIL population described here, from 3 back-cross generations from the same parents, and from the literature were earlier analyzed in a meta analysis (Lacape et al. 2010). A small number (26) of regions of QTL coincidence, so-called meta-clusters, of phQTLs were further delineated by a 10–20 cM (except for FIN_21A framed by 27 cM) window on the RIL map, depending on the support intervals of individual QTLs within the clusters. The overlap between these phQTL meta-clusters and eQTLs was assessed by the direct comparison of their genomic distribution on the RIL map and using the one-LOD drop off confidence intervals of the eQTLs. Comparative distribution of phQTL meta-clusters and eQTLs along the 26 chromosomes are shown in Online Resource ESM3 (2 chromosomes, c2 and c21 shown in Fig. 3 as examples).

There was no bias in the localizations of the eQTLs with higher LOD values (10% of LOD > 5) from the 10 and 22 dpa experiments as compared to the meta-clusters of phQTLs (not shown). Because TDFs are anonymous and because eQTLs are widely distributed throughout the genome (only 20% of all bins mapped no eQTL), we concentrated on the co-localization of meta-clusters of phQTLs and the hotspots of eQTLs, rather than the individual eQTLs themselves. In each data set, the density of eQTLs within meta-clusters of phQTLs was significantly higher than random (Chi^2 probability <E-10 and <E-12 for the 10 and 22 dpa data sets). In several instances, the meta-clusters of fiber QTLs overlapped with regions significantly enriched in eQTLs (hotspots) for either, or both, the 10 and 22 dpa samples. Interestingly, the region most highly populated in eQTLs for the 10 dpa samples along a central region of c21 corresponded fairly well with two co-localized meta-clusters of phenotypic QTLs (Fig. 3), for fiber strength (STR_21) and fineness (FIN_21A). For both traits, the Gb parent contributed positively (higher strength, lower fineness). Including c21, coincident enrichment in eQTLs (hotspots) within phQTL meta-clusters were encountered in nine cases. Three involved both 10 and 22 dpa eQTL hotspots (c2 containing a meta-cluster for fiber elongation, c9 and c21-lower for fineness), four involved only the 10 dpa eQTLs (c3 and c21-upper for fineness and strength, c4 for length, and c12 for fineness) and two involved only the 22 dpa eQTLs (c17 for fineness and c24 for length) (Table 3; Fig. 3, Online Resource ESM3).

Two chromosomes particularly rich in phQTLs clusters that were only reported as indicative in Lacape et al. (2010) because of imperfect clustering and which co-localized with hotspots of eQTLs, were c5 and c19. An upper region of c5 rich in fiber fineness and strength phQTLs coincided with eQTL hotspots detected in both 10 and 22 dpa along bins 11–14. On chromosome 19, although eQTL hotspots
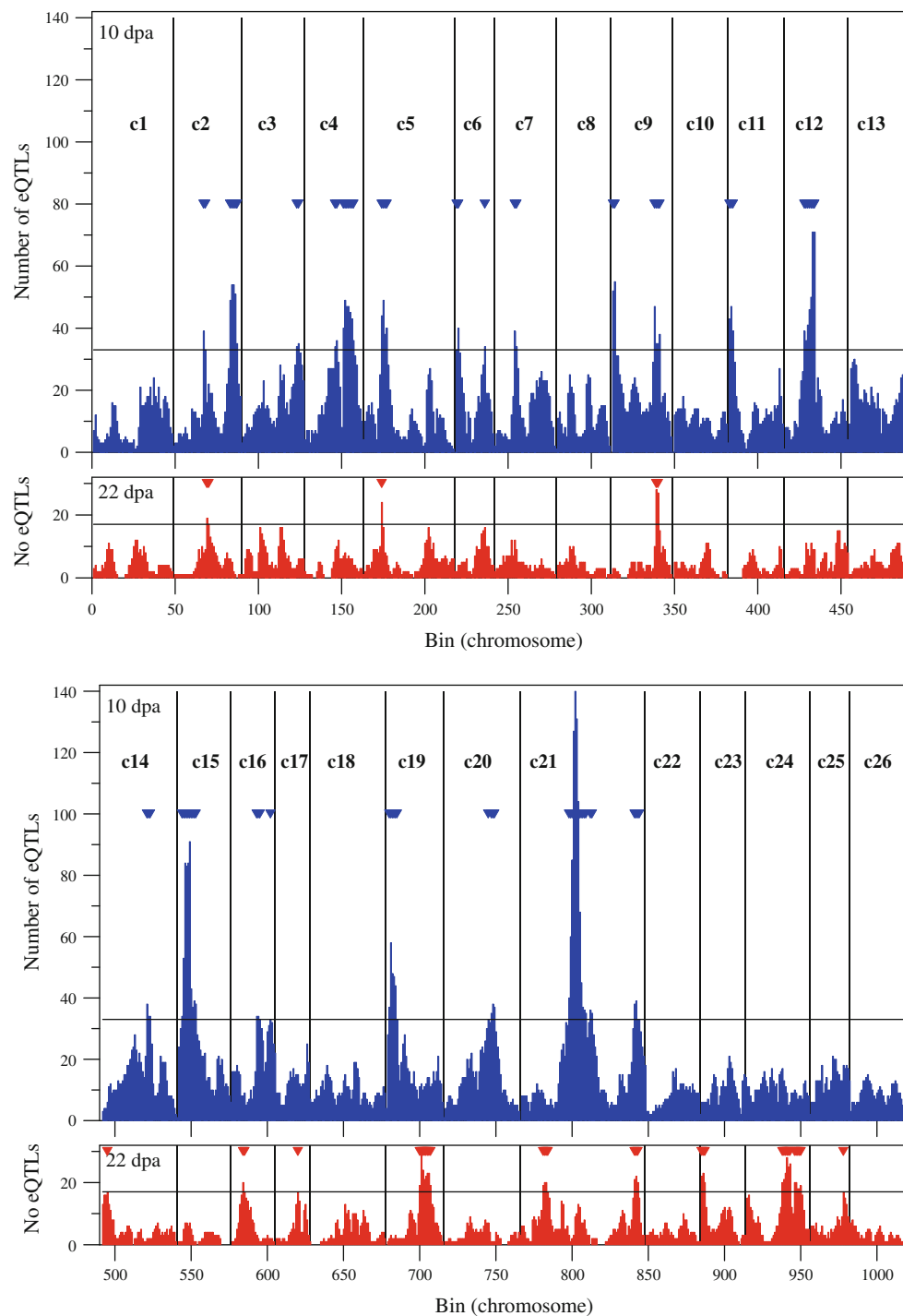
Fig. 2 Localization of eQTLs along chromosomes. Distribution of significant (LOD > 3.5) eQTLs along chromosomes of the A (c1–c13) and D (c14–c26) sub-genomes of tetraploid cotton according to bins of 2 cM of the Guazuncho 2× VH8-RIL map. eQTLs detected from fibers of two different developmental ages are presented: 10 dpa (blue) in the upper panels and 22 dpa (red) in the lower panels. The *horizontal lines* represent the 1,000 permutation-based ($P < 0.05$) thresholds, as 33 and 17 eQTLs per bin in the 10 and 22 dpa experiments, respectively. Individual bins with total eQTLs higher than threshold are indicated (*arrow heads*) (color figure online)

were detected at different locations in each 10 dpa (along bins 2–7) and 22 dpa (bins 22–29) series, none were co-located with the putative fiber phQTL clusters for elongation and fineness that occur on those chromosomes.

Quantitative RT-PCR

As AFLP bands are anonymous, 29 selected TDF fragments were isolated from gels and sequenced to tentatively

Table 3 Summary data of eQTL hotspots in the 10 and 22 dpa experiments

Experiment	Common 10–22	Chr	Bin number	SSR loci in interval (RIL map)			No. 2 cM bins	No. eQTL peaks	No. overlap. CI	Mean LOD	Mean R²	Add Gb^*	Add Gh^*	Chi²	Co-localized phQTL meta-cluster
Experiment: 10 dpa															
10 dpa	×	2	c2.18/19	MUCS399a	BNL1413	JESPR101b	2	42	72	5.0	0.18	25	17	0.47 NS	ELO_2 (Gh), FIN_2 (Gh)
10 dpa		2	c2.34/38	NAU5384	CIR376c	NAU1246	5	72	243	4.4	0.16	42	30	0.46 NS	
10 dpa		3	c3.33/34	CIR263	BNL3259a	NAU2161	2	41	69	4.2	0.16	27	14	0.13 NS	FIN_3 (Gb), STR_3 (Gb)
10 dpa		4	c4.18/19	CIR048a	CIR291b	BNL4047	2	41	70	6.6	0.21	26	15	0.23 NS	
10 dpa		4	c4.23/29	CIR222b	NAU2231b	CIR187c	7	77	307	4.7	0.21	40	37	0.72 NS	LEN_4 (Gh)
10 dpa	×	5	c5.11/14	BNL4071a	NAU2029	CIR062a	4	71	171	4.5	0.17	39	32	0.88 NS	(putative FIN (Gh) and LEN (Gh))
10 dpa		6	c6.01/02	CIR203		CIR017	2	40	73	4.9	0.18	24	16	0.45 NS	
10 dpa		6	c6.18	CIR291f	BNL2691b	BNL2884	1	34	34	4.6	0.16	20	14	0.57 NS	
10 dpa		7	c7.12/13	JESP228b		CIR323	2	43	73	4.7	0.19	26	17	0.39 NS	
10 dpa		9	c9.01/02		BNL0686a		2	55	107	4.4	0.19	28	27	0.65 NS	(FIN_9A (Gb))
10 dpa	×	9	c9.26/29	MUSS547	BNL3582	MUSS083	4	90	155	4.7	0.17	50	40	0.77 NS	FIN_9B (Gb)
10 dpa		11	c11.01/03	CIR196	CIR003	BNL1066	3	53	129	4.7	0.17	28	25	0.86 NS	
10 dpa		12	c12.12/18	BNL3537b	BNL1673	CIR167	7	124	354	4.9	0.18	57	67	0.07 NS	FIN_12 (Gb)
10 dpa		14	c14.30/32	JESPR165	CIR228b	BNL0226a	3	45	106	4.8	0.18	25	20	0.83 NS	
10 dpa		15	c15.03/12	MUSS161a	MGHES332	MUSB0818c	10	125	586	5.1	0.19	65	60	0.65 NS	
10 dpa		16	c16.17/19	NAU2432a	CIR100	MGHES75	3	44	101	4.6	0.17	32	12	0.01*	
10 dpa		16	c16.26	CIR107		NAU1020b	1	33	33	5.1	0.20	19	14	0.68 NS	
10 dpa		19	c19.02/07	CIR224b	BNL3452	CIR176	6	76	267	4.8	0.18	34	42	0.11 NS	
10 dpa		20	c20.29/33[b]	JESPR244a	BNL3071	JESPR235a	5	68	175	4.7	0.17	27	41	0.02*	
10 dpa		21	c21.32/47[b]	BNL1551a	CIR061c	BNL1053b	16	250	1,038	5.1	0.19	155	95	0.01*	FIN_21A (Gb), STR_21 (Gb)
10 dpa	×	21	c21.75/78[b]	BNL3147b	CIR013	CIR254b	4	44	143	4.7	0.18	22	22	0.59 NS	FIN_21B (Gb)
		Tot.	(21)				91	1,468	4,306	4.9	0.18				

Experiment	Common 10–22	Chr	Bin number	SSR loci in interval (RIL map)			No. 2 cM bins	No. eQTL peaks	No. overlap. CI	Mean LOD	Mean R²	Add Gb^a	Add Gh^a	Chi²	co-localized phQTL meta-cluster
Experiment: 22 dpa															
22 dpa	×	2	c2.20/21	MUCS399a	JESPR101b	NAU3875	2	20	36	4.5	0.15	6	14	0.06 NS	ELO_2 (Gh), FIN_2 (Gh)
22 dpa	×	5	c5.11	BNL4071a	BNL3029a	NAU2029	1	24	24	4.9	0.17	13	11	0.76 NS	
22 dpa	×	9	c9.27/28	NAU3365	JESPR248b	BNL1317b	2	35	55	4.7	0.16	13	22	0.10 NS	FIN_9B (Gb)
22 dpa		14	c14.04	CIR246	CIR246		1	17	17	4.5	0.18	9	8	0.87 NS	
22 dpa		16	c16.08/09	Coau4E21	BNL1122b	JESPR228a	2	22	37	4.6	0.16	9	13	0.34 NS	
22 dpa		17	c17.15	JESPR195	JESPR101a	JESPR101a	1	17	17	4.5	0.16	13	4	0.04*	FIN_17 (Gb)
22 dpa		19	c19.22/29	CIR062b	CIR179b	BNL3347b	8	61	182	4.9	0.17	23	38	0.04*	
22 dpa		21	c21.15/18	BNL3279		BNL0836b	4	24	76	5.2	0.19	14	10	0.47 NS	
22 dpa	×	21	c21.75/77	NAU1148b	CIR254b	CIR254b	3	23	63	4.9	0.18	8	15	0.12 NS	FIN_21B (Gb)
22 dpa		23	c23.01/03	NAU1035	CIR286	NAU3214a	3	23	65	4.2	0.15	8	15	0.12 NS	

Table 3 continued

Experiment	Common 10–22	Chr	Bin number	SSR loci in interval (RIL map)			No. 2 cM bins	No. eQTL peaks	No. overlap. CI	Mean LOD	Mean R²	Add Gbᵃ	Add Ghᵃ	Chi²	co-localized phQTL meta-cluster
22 dpa		24	c24.24/36ᶜ	MUSS500	BNL0387b	BNL0252	13	83	260	4.8	0.17	39	44	0.46 NS	LEN_24 (Gb)
22 dpa		25	c25.22	BNL1440b	CIR407	BNL1417	1	17	17	4.6	0.17	9	8	0.87 NS	
	Tot.	(12)					41	366	849	4.8	0.17				

Bin (2 cM long) numbers and locus names are referring to the RIL map, as reported in Lacape et al. (2009)

* Significant (P < 0.05) Chi² values indicating unbalanced proportion of parental additivity effects among eQTLs within eQTL hotspot

ᵃ No. eQTLs with additive effect from Gb or Gh parent

ᵇ Bins 20.30, 21.44, and 21.45, with 30, 27, and 32 eQTLs are slightly below 10 dpa threshold (33 eQTLs/bin)

ᶜ Bin 24.30 and 24.31, with 12 and 10 eQTLs, are below 22 dpa threshold (17 eQTLs/bin)

confirm by qPCR the population-wide quantitative cDNA-AFLP profiling. Ten TDFs were selected based on the following criteria: (1) the correlation of the TDF expression data with the fiber phenotype (length, strength, or fineness), and (2) the coincident chromosome localization of the TDF-derived eQTL and fiber trait-derived phQTLs. In addition, 19 other TDFs were selected based on the high LOD score (LOD > 5) of at least one of their associated eQTL(s). The sequences of these 29 TDFs were aligned against the cotton sequence databases using a Blast algorithm. Out of the 29 TDFs, 15 were subsequently tested by qPCR (Online Resource ESM2). Sequence homology of the TDFs with cotton EST sequences allowed the design of gene-specific primers. Expression profiling over the same 88 RILs was carried out using qPCR. For three TDFs, for which the sequence blast result could not discriminate between several possible database accessions, primers pairs were designed for each accession and tested independently by qPCR (Online Resource ESM2).

Gene expression qPCR values over the 88 RILs were analyzed by WinQTL Cartographer. Those gene expression data sets that deviated from a normal distribution were log-transformed before analysis. LOD peak values were in general lower for the qPCR data so we used LOD2.5 as a threshold for the comparisons with the cDNA-AFLP eQTLs. All candidate genes tested mapped at least one eQTL (>LOD2.5). The comparison of the position of the qPCR-derived eQTLs with the AFLP-derived eQTLs of the corresponding TDF showed six cases of a common location of at least one LOD peak (position <10 cM) on the same chromosome. However, for half of those cases, the additive effects were in opposite directions. The only three congruent cases are shown in Fig. 4 (LOD and additivity profiles). The first TDF (Fig. 4a), BCTMGA_95.6, putatively encoding a UDP-glucose pyrophosphorylase (accession DT050294), tested by qPCR mapped an eQTL (LOD2.5) on c24 at a similar location as a cDNA-AFLP peak of lower LOD value (LOD1.2), while conversely an AFLP-based eQTL on c5 (LOD3) was corroborated by a lower LOD peak in the qPCR-based eQTL mapping (LOD1.1). The profiles of the additive effects along c5 and c24 were fairly similar for both techniques (Fig. 4a). The second TDF, BCGMCT_399, had a Blast hit with a D-ribulose-5-phosphate 3-epimerase (accession EX167 888), and showed congruent eQTL localization in a middle region of c21 (Fig. 4b). An additional LOD peak (LOD2.5) on the same chromosome detected by the qPCR was not detected in the AFLP experiment. The third TDF, BCTMGA_133.9, had a blast hit with an unknown protein (EX166831), and showed congruent LOD and additivity profiles (Fig. 4c) on two chromosomes, c5 and c21; although, as for the previous TDF, some LOD peaks were not clearly confirmed with the two methods.
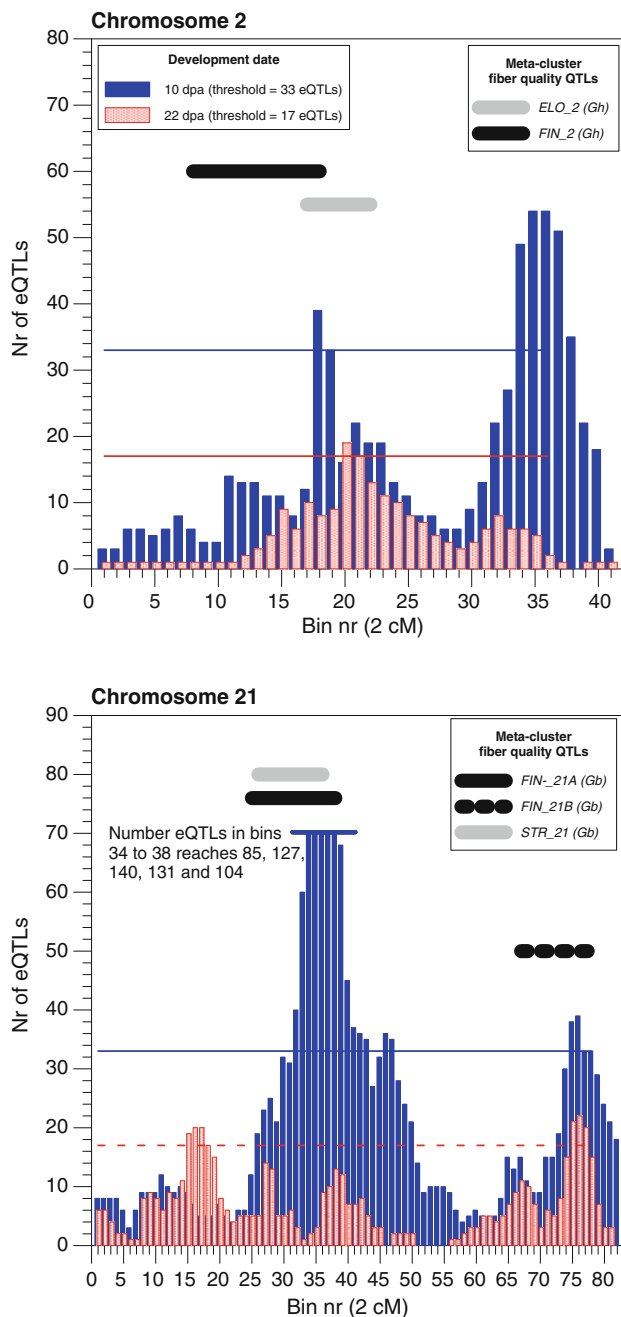
Fig. 3 Comparative localization of eQTLs with meta-clusters for fiber phenotypic QTLs. Localization of eQTLs from fibers at 10 (*blue bars*) and 22 dpa (*red bars*) are compared with regions containing meta-clusters of fiber phenotypic QTLs (phQTLs) for major fiber quality trait categories, including fineness (FIN), length (LEN), strength (STR), elongation (ELO) and color (COL) as reported in Lacape et al. (2010). The observed parental effect (either *Gh* or *Gb*) of each meta-cluster of phQTLs, as indicated in the legends, corresponds to an improvement of the trait value, higher length, strength, and elongation or lower fineness and yellowness. The horizontal lines represent the 1,000 permutation-based (*P* < 0.05) thresholds in the 10 and 22 dpa experiments. Examples of chromosomes c2 and c21 are shown (complete set of 26 chromosomes are shown in Online Resource ESM3) (color figure online)
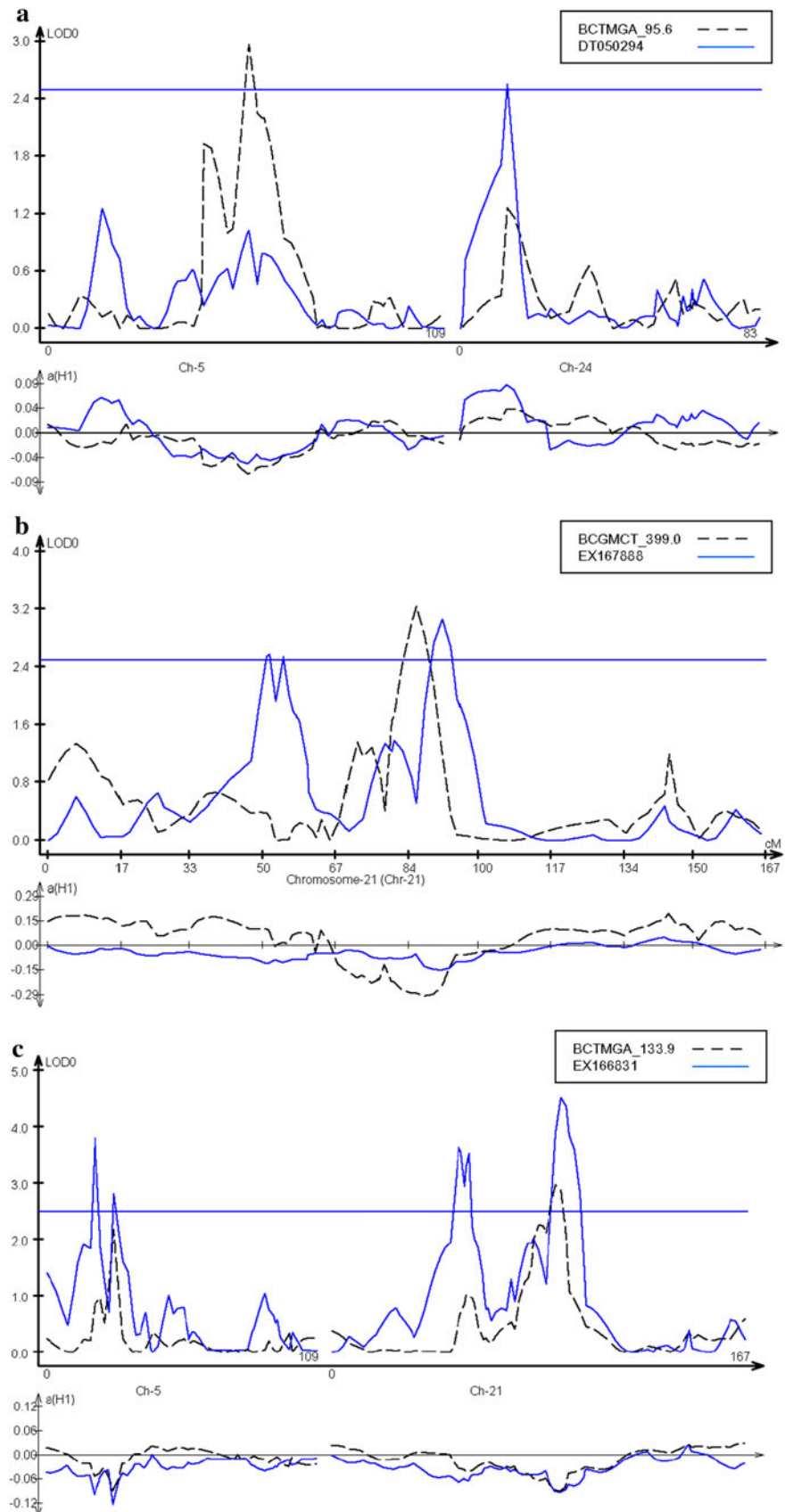
Linking cDNA-AFLP and EST sequence data

The AFLP in silico digestion of the 38,297 contigs from the global EST assembly (unpublished) resulted in 11,421 predicted TDFs (a unigene predicts a single TDF) of size range between 6 and 2,150 nt. Because of the lower and upper limits imposed by the DNA ladder used, the search for matches was limited to the range 50–340 nt that were scored on our cDNA-AFLP gels. As a result, subsets of 1,970 and 661 cDNA-AFLP TDFs, were selected among the 3,665 and 1,375 TDFs mapped in the two eQTL experiments, 10 and 22 dpa, respectively. Similarly 8,087 pTDFs in the range 50–340 bp were used for comparisons. The search for matches of pTDFs with TDFs quantified on the cDNA-AFLP gels indicated that 1,237 and 448 TDFs (or 63 and 68%) in the 10 and 22 dpa eQTL experiments, respectively, could be predicted by one or several annotated unigenes. The AFLP in silico analysis of a global transcriptome assembly constitutes a valuable tool (Qin et al. 2006) that could be used to identify digitally interesting expression patterns, or used in a preliminary step of mining of pre-existing EST data in order to optimize choice(s) of enzymes in cDNA-AFLP studies (Rombauts et al. 2003; Vuylsteke et al. 2007).

Among the 265 unigenes with significant differential digital expression (fold change larger than 2) between the 2 genotypes at 10 dpa (not shown) and which predicted a TDF in silico, 167 corresponded to a TDF mapping at least one eQTL among the RILs at 10 dpa. A similar number of 265 unigenes, although the 2 lists are essentially different, predicting a TDF in silico showed significant differential digital expression between the 2 genotypes at 22 dpa (not shown); however, only 64 of them mapped at least one eQTL among the RILs at 22 dpa. All these 227 unigenes and their putative annotations by Blast2Go are listed in Online Resource ESM4. Among the interesting annotations that were detected in this list, some proteins are known for their pivotal role in cotton fiber biogenesis and elongation and in explaining differences between the two species *G. hirsutum* and *G. barbadense*, like kinases, beta galactosidase, cellulose synthases, tubulins, or endo-beta glucanases (Online Resource ESM4).

Discussion

We here report a cotton fiber transcriptome study by cDNA-AFLP in a segregating RIL population of interspecific origin (*G. hirsutum* × *G. barbadense*), and thus have provided the first demonstration of an eQTL analysis in cotton.

Fig. 4 LOD and additivity profiles from cDNA-AFLP and qPCR experiments. The three examples of comparative LOD (*upper panels*) and additivity (*lower panels*) profiles derive from gene and TDF expression data as profiled in quantitative RT-PCR and cDNA-AFLP experiments. *Dotted line* TDF profile from cDNA-AFLP, *solid line* gene profile from qRT-PCR. *Horizontal line* indicates a LOD2.5 threshold. Examples show cases of co-localizations of LOD peaks and/or parallel profiles:
a TDF = BCTMGA1_95.6 and gene DT050294 on c5 and c24, **b** BCGMCT_399 and gene T19 on c21; **c** BCTMGA_133.9 and gene FT6-unk on c5 and c21 (color figure online)

Quantitative cDNA-AFLP is efficient for gene expression profiling

While the small population size (88 individuals) used in this study and the unbalanced parental representation (distorted segregation) present in these RILs (Lacape et al. 2009), could be viewed as limitations (Beavis 1998), the use of environmentally controlled growth facilities, standardized times of collection of samples and a relative stringent QTL detection threshold are all factors in favor of increased rigor in the identification of eQTLs. Our LOD threshold, LOD3.5 (equivalent to a 5% genome-wide risk) was fairly stringent when compared to similar eQTL studies [LOD2.6 in West et al. (2007), or LOD2.9 in Potokina et al. (2008), and LOD3.1 in Wang et al. (2010)]. Gene expression may also be influenced by the environment. Although our experiment did not include separate biological replicates as such, our sampling (each RNA originating from several bolls harvested from two different plants), may have at least partly minimized this environmental component. We expected that the eQTLs detected through our design would be more related to differences in fiber development processes than to variation of environmental conditions and genetic background.

Two key fiber development time-points were studied, whereby RNA of developing fibers at 10 and 22 days after anthesis were profiled over the segregating RILs. 4,464 transcripts were profiled, thus representing only a modest, 10–14%, part of the total fiber transcriptome [75–94% of the 40,000 genes of the genome are expressed in cotton fibers (Hovav et al. 2008c)]. Using a 5% genome-wide permutation-based threshold and a global risk of 5% (LOD > 3.5) 5,040 significant eQTLs at the two key times of fiber development were identified, deriving from 3,023 different TDFs. As expected from previous classical fiber QTL mapping studies and from functional genomics work, the genetical genomics approach and the eQTL networks of developing fibers identified here confirmed the complexity of the transcriptional regulation leading to the contrasting fiber qualities of the two species G. hirsutum and G. barbadense. The majority of the eQTLs were associated with relatively small effects: only 7 and 3% of the eQTLs had an $R^2 > 0.3$ in the 10 and 22 experiments, respectively. This is similar to findings in yeast and other higher plant species for expression traits (Keurenjes et al. 2007; West et al. 2007). This also implies that most of the gene expression variation in our experiment may be linked to interactions with the environment, the particularly heterogeneous genetic background of those RILs, or to non-controlled technical bias. Although all RILs were grown in the same glasshouse environment, considerable variability in their growth (including fruit maturation kinetics) and development (including earliness and possible delays in fruit and fiber maturation) was observed (not shown).

The cDNA-AFLP profiles of 10 and 22 dpa fibers were different in terms of the total number of scored TDFs (3,962 against 1,404), although they had very similar basic statistics: including information content (two-third of all TDFs mapped at least 1 eQTL), similar mean R^2 and additivity. The main reason for the difference may be technical as there was a greater difficulty in choosing quantitatively segregating bands in the 22 dpa radiograms that had more "background noise". RNA from older (22 dpa) fibers was generally lower in quantity and quality and those fibers may also be less synchronous material in physiological development compared to younger (10 dpa) fibers. Boll size, boll locks and seed size observed at 22 dpa did display more variability than at the 10 dpa sampling date (although not formally measured).

Sub-genomic distribution of eQTLs

Quantitative trait loci mapping reports for fiber traits in tetraploid cotton have indicated that the contribution of the D_t sub-genome to fiber quality is far from negligible (Lacape et al. 2010; Rong et al. 2007), despite the fact that the modern diploid species from the D sub-genome produce only very rudimentary short fibers compared to the long fibers of the A and $A_t D_t$ species. The numbers of eQTLs mapping to the A_t (1,925) and D_t (1,740) chromosomes of tetraploid cotton in this study were quite similar. However, the distribution of the eQTL hotspots between A_t and D_t chromosomes as identified from the 22 dpa fibers seemed to favor the D_t chromosomes (3 hotspots vs. 9 for the A_t and D_t chromosomes, respectively), while the tendency was opposite for the 10 dpa hotpots (13 vs. 8, respectively). This agrees with the observations of Hovav et al. (2008b) that noted a preference for expression of the D_t genome copies within homoeologous gene pairs compared to the A_t genome, and this difference increased during fiber development. Hovav et al. (2008b) used a microarray platform of 1,500 duplicated A_t/D_t gene pairs and showed that changes in the expression levels of duplicated A_t/D_t genes, including their temporal differential partitioning, was a common phenomenon. That neither phQTLs [or phQTL meta-cluster as in Lacape et al. (2010)] nor eQTLs (or eQTL hotspots, herein) were detected at homoeologous locations on A_t and D_t chromosomes is in favor of a decoupling (i.e. lack of coordination) of the expression of homoeologous duplicated genes, or gene regulators, in fibers of tetraploid $A_t D_t$ cotton.

Existence of hotspots of eQTLs

eQTLs can be considered as either cis-acting or trans-acting types according to their distance to the corresponding structural gene (Hansen et al. 2007). Although

trans-eQTL hotspots have emerged as an important aspect of genetic architecture of transcription and interpreted by the location of a transcription regulator with pleiotropic effects on a large set of genes (Gibson and Weir 2005; Kirst et al. 2005), the interpretation of the clustering of eQTLs at a particular hotspot should at the same time be considered with some caution. In the absence of a genome sequence for cotton, we are as yet unable to clearly define which of our eQTLs are *cis* or *trans*, although the occurrence of highly populated hotspots is clearly suggestive of *trans* eQTLs. Another interpretation is that they may reflect regions of the map with little recombination, and hence may have more genes per cM than elsewhere, as had been reported in the case of barley (Potokina et al. 2008). In our case, we found no relationship between the map position of eQTL hotspots and the position of marker-dense regions in the RIL map (Lacape et al. 2009). Another factor leading to an over-estimation of *trans*-eQTL occurrence could be artificially or environmentally induced inter-sample correlations (Kang et al. 2008; Pérez-Enciso et al. 2007), In such cases, the existence of false eQTL hotspots generated by highly correlated genes can be assessed using a permutation strategy (Breitling et al. 2008), that was applied to the 10 and 22 dpa data matrix. The genotypic data of the 88 individual RILs were permuted maintaining the expression matrix and an SMA was re-run with QTLCartographer. Using $\Pr(F) < 0.001$ as a threshold to filter the highest gene × marker correlations, and following two runs of permutations for each of the 10 and 22 dpa data sets, it was found that none of the individual hotspot loci (25 loci in the 10 dpa data and 9 loci in the 22 dpa data) in the real data (observed hotspots) was among the hotspot loci in the two permuted data sets (false hotspots). The observed eQTL hotspots identified in this study (Table 3) are therefore unlikely to be explained by simple gene expression correlations and may represent true *trans* eQTLs.

Our results on the mean statistics amongst the eQTLs mapped within hotspots as compared to all eQTLs (mean R^2 and LOD value for hotspots and bias of parental additive effects) differ from the literature. The most common observation for eQTL hotspots is that (1) significantly lower LOD scores and lower R^2 values are associated with *trans*-eQTLs hotspots (Gibson and Weir 2005; Potokina et al. 2008; Vuylsteke et al. 2006; West et al. 2007), although we observed very similar values (Table 3), and (2) a directional bias of additive effects for trans-eQTL hotspots was strong for Arabidopsis (West et al. 2007), and moderate for eucalyptus (Kirst et al. 2005) or barley (Potokina et al. 2008), but we observed only a slight bias in parental additive effects (Table 3).

The distributions of our eQTL hotspots were compared with the position of fiber "gene-rich islands" identified in Xu et al. (2008). They identified a strong and significant bias in the map location of fiber gene-derived markers in favor of a few chromosome regions, specifically on c5, c10, c14 and c15. Among those 4 regions, a 10 cM region on top of c5 mapped 460 genes and putatively corresponded to the region of the RIL map (bins 5_11 to 5_14) where an eQTL hotspot common to the 10 and 22 dpa data sets occurs (Online Resource ESM3). In this particular region of c5, the eQTL hotspots may therefore reflect the presence of a gene-rich region rather than of a pleiotropic *trans*-acting factor. It should be noted that the genes listed by Xu et al. (2008), although expressed in fibers, are not necessarily differentially expressed between *G. hirsutum* and *G. barbadense*. Those genes will not necessarily underlie or be causal for the strong phenotypic QTLs clusters earlier described (Lacape et al. 2010) or the eQTLs targeted by the cDNA-AFLP method described herein.

Integration of fiber phQTLs and eQTLs

Important practical applications of global scanning of the variation in transcript abundance and eQTL mapping includes the correlative analysis of trait variation and the comparative mapping of eQTLs with phQTLs.

Because of the lack of sequence data and because of the limited number of cDNA-AFLP-derived transcripts profiled in this study, our data could not be used to identify *cis*-acting factors putatively influential in the variation of phenotypic traits. A concordant localization of eQTL hotspots (presumably associated with *trans*-acting factors) with regions rich in phenotypic fiber QTL (meta-clusters) was observed in at least 15 different cases related to the various fiber traits (Table 3; Fig. 3 and Online Resource ESM3) the majority of which related to fiber fineness and to a lesser extent to fiber elongation, strength and length. It is generally accepted that among the main fiber quality parameters, fiber length and fineness are determined in the early stages of growth and development, while strength–stage relationships need to be further studied (Hsieh 1999). However, there was no clear correspondence of eQTL hotspots detected for the two fiber development stages with any specific type of fiber quality parameter, in terms of co-localization with meta-clusters of phQTLs (Table 3). An eQTL may therefore involve pleiotropic factors or modulate in an unexpected way the output of physiological processes involved in fiber development (e.g. length can be modified by an extended elongation phase, and fiber strength can also depend on early structures of fiber cells that might determine fiber diameter).

Other studies have sometimes seen a co-localization between eQTLs and phQTLs, but this is not always the case. In Arabidopsis, a summation approach combining phenotypic QTLs from unrelated studies (62 traits) did not reveal any significant co-localization with *trans*-eQTL

hotspots (Kliebenstein et al. 2006), except for two shoot growth QTLs. In barley, a number of comparative mapping studies in relation to rust resistance concluded that there was either an association or an independence between specific traits and some eQTLs (Chen et al. 2010; Druka et al. 2008). In an Eucalyptus hybrid progeny, two *trans*-acting eQTLs common to 11 genes from the lignin biosynthesis pathways were shown to be co-localized with growth phQTLs (Kirst et al. 2004).

Our study may have been an unusual case of eQTL mapping since we mapped expression polymorphisms between two independently domesticated tetraploid species rather than between different cultivars or ecotypes of the same species normally used in genetic studies. Molecular studies on the domestication of *G. hirsutum* and *G. barbadense* have already suggested that human selection has acted on different components of the fiber developmental genetic program and that convergent evolution rather than parallel evolution has affected fiber traits in both species (Hovav et al. 2008a).

Nevertheless, the genomic co-localization of phQTL and eQTL hotspots that we observed may also partly reflect the consequences of domestication. Sequences data from cloning of domestication-related genes and diversification-related genes strongly suggest that domestication is associated with changes in transcriptional regulatory networks, whereas crop diversification involves a larger proportion of enzyme-encoding loci (Doebley et al. 2006). As *Gh* and *Gb* have experienced separate domestications the RILs should segregate for domestication related traits. As domestication-related genes may be transcription factors and act in *trans*, some of the highlighted regions showing co-localization of eQTL hotspots and phQTL meta-clusters may therefore reflect the genomic consequences of domestication in both species (e.g. c3, c21 or c24 for *Gb*, c2 or c4 for *Gh*) (Fig. 3 and Online Resource ESM3).

The few TDFs showing higher R^2 values, good correlations with fiber trait parameters and localized within phQTL hotspots will be a useful starting point to investigate the importance of *trans*- versus *cis*- regulations in the determination of fiber trait parameters. Such sequences could also constitute good candidate genes that have been targeted by human selection in each of the two species. With the future sequencing of the cotton D genome (Paterson et al. 2010), map-based candidate sequences for selection underlying phQTL or eQTL should also be available. As a first example, a R2R3-MYB transcription factor showing the highest number of haplotypes (in four *Gossypium* species) among 12 MYB genes studied has been mapped on c21 (An et al. 2008) near the marker JESPR251, i.e., in the center of where we mapped an important hotspot of eQTLs as well as a cluster of fiber phenotypic QTLs (Fig. 3).

Validation of TDFs

Because of the large number of eQTLs identified, validation was only undertaken for a small set of 15 cloned TDFs that were tested by quantitative RT-PCR across the same RILs. The eQTLs derived from the 2 approaches (cDNA-AFLP and qPCR) were only partly coincident, as only 3 clear cases of congruence were detected among the 15 examined. To our knowledge, there have been no previous reports in plants of validation by quantitative RT-PCR of eQTLs mapped from microarray or from AFLP profiling. Only recently, in a report on eQTL mapped from microarray, 13 genes from porcine muscle were tentatively validated by qPCR (Ponsuksili et al. 2010). As a result, 3 of the 5 *cis*-acting eQTL were validated and none of the 15 *trans*-eQTLs was commonly detected in the 2 methods. Possible explanations for our modest level of validation may be several. The isolation of the PCR fragments from AFLP gels (used for cloning the TDF) is technically difficult and may have selected a different band of the same or similar size from that identified in the original gels used for eQTL mapping. Alternatively the choice of primers for qPCR may have amplified a transcript for a different gene to the one corresponding to the TDF. The Genbank EST collection, while extensive, is by no means complete and *Gb* ESTs in particular are significantly under-represented. After Blasting our sequences against Genbank ESTs, primers were designed to the 3′ UTR region, sometimes outside the sequence region overlapping with the TDF and, so could possibly correspond to a different (homoeologous, paralogous) gene copy. The three cases (Fig. 4), where congruent profiles were found for LOD position and additivity, could not happen by chance and constitute a validation of cDNA-AFLP as a technique to quantitatively monitor the segregation of transcript abundance across a population (Vuylsteke et al. 2006).

The cross-matching between virtual TDFs predicted by in silico AFLP digestion of a EST unigene and AFLP-based TDFs allowed the identification of 227 cases where an annotated unigene, with differential digital representation of the 2 parents, predicted a TDF matched with a AFLP-based TDF fragment also mapping an eQTL for the same developmental time point (Online Resource ESM4). These annotated unigenes are all candidate sequences for validation of the cDNA-AFLP technique. The identified unigenes could be screened by qRT-PCR for verification of quantitative changes in transcript abundance (gene expression) both for the cDNA-AFLP and for the digital estimates from EST sequence assembly. The positioning of these genes on the cotton genome, when available (Lin et al. 2010; Paterson et al. 2010), will provide an aid for the interpretation of *cis*- (same position of the gene and of the eQTL) and *trans*-acting (different positions) eQTLs.

Conclusion

In conclusion, we have not only validated the concept of genetical genomics (Jansen and Nap 2001) for the first time in cotton but also demonstrated that cDNA-AFLP is a cost-effective and highly transferable alternative for genome-wide and population-wide gene expression profiling (Reijans et al. 2003). The cDNA-AFLP technique can thus be used in a cost-effective way in any biological system. While it is not practical to sequence all of the TDFs we have mapped, their sequence identity could be determined through in silico prediction (Rombauts et al. 2003) using the increasing number of *Gossypium* cDNA sequence data. For that purpose and as a continuation of the present report, we have undertaken a series of large-scale sequencing of cDNA from the same material, including the two parental species and the two fiber development time points. As the genome sequence of the D species *G. raimondii* should be available in the next few years (Paterson et al. 2010), the identification of candidate genes underlying phQTLs or eQTLs and the classification of eQTL as *cis*- or *trans*- eQTLs will also be facilitated (Wang et al. 2010).

The identification of putative *trans*-eQTL hotspots controlling large numbers of fiber transcripts has been an important output and has been observed in many eQTLs studies in other organisms and plants (Brem et al. 2002; Schadt et al. 2003; West et al. 2007). How those eQTL hotspots may highlight effects of human selection on plant genomes is an exciting question raised by this study. Candidate genes could then be tested as fiber quality markers in *G. hirsutum* and *G. barbadense* collections, including their native wild and domesticated forms (Zhu et al. 2008) or in introgressed near-isogenic lines, or tested for footprints of domestication in *Gossypium* species. Finally, a better knowledge about the molecular control of cotton fiber quality will allow breeders to choose better genomic targets for improvement of this crop species through marker-assisted selection and genotypic construction.

References

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized allotetraploid. Genetics 168:2217–2226

Al-Ghazi Y, Boutrot S, Arioli T, Dennis ES, Llewellyn D (2009) Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. Plant Cell Physiol 50:1364–1381

Alabady M, Youn E, Wilkins TA (2008) Double feature selection and cluster analyses in mining of microarray data from cotton. BMC Genomics 9:295

Albertini E, Marconi G, Barcaccia G, Raggi L, Falcinelli M (2004) Isolation of candidate genes for apomixis in *Poa pratensis* L. Plant Mol Biol 56:879–894

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

An C, Saha S, Jenkins JN, Ma DP, Scheffler B, Kohel RJ, Yu J, Stelly DM (2008) Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome, and linkage mapping. Theor Appl Genet 116:1015–1026

Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci AM, Da Silva C, Cascardo J, Allegre M, Kuhn D, Verica J, Courtois B, Loor G, Babin R, Sounigo O, Ducamp M, Guiltinan MJ, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. BMC Genomics 9:512

Bachem CW, van de Hoevan RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9:745–753

Basra AS, Malik CP (1984) Development of the cotton fiber. Int Rev Cytol 89:65–113

Basten CJ, Weir BS, Zeng ZB (2003) QTL Cartographer Version 1.17: a reference manual and tutorial for QTL mapping. Department of Statistics. North Carolina State University, Raleigh

Beavis WD (1998) QTL analyses: power, precision and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, New York, pp 145–162

Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, Gerrits A, Bystrykh LV, de Haan G, Su AI, Jansen RC (2008) Genetical genomics: spotlight on QTL hotspots. PLoS Genet 4:e1000232

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Nat Genet 296:752–754

Breyne P, Dreesen R, Cannoot B, Rombaut D, Vandepoele K, Rombauts S, Vanderhaeghen R, Inzé Z, Zabeau M (2003) Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Genet Genomics 269:173–179

Chen X, Hackett CA, Niks RE, Hedley PE, Booth C, Druka A, Marcel TC, Vels A, Bayer M, Milne I, Morris J, Ramsay L, Marshall D, Cardle L, Waugh R (2010) An eQTL analysis of partial resistance to *Puccinia hordei* in barley. PLoS One 5:e8598

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321

Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W (2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol 48:75–97

Druka A, Potokina E, Luo Z, Bonar N, Druka I, Zhang L, Marshall DF, Steffenson BJ, Close TJ, Wise RP, Kleinhofs A, Williams RW, Kearsey MJ, Waugh R (2008) Exploiting regulatory variation to identify genes underlying quantitative resistance to the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* in barley. Theor Appl Genet 117:261–272

Gibson G, Weir B (2005) The quantitative genetics of transcription. Trends Genet 21:616–623

Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 24:408–415

Guo S, Zheng Y, Joung JG, Liu S, Zhang Z, Crasta OR, Sobral BW, Xu Y, Huang S, Fei Z (2010) Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. BMC Genomics 11:384

Hansen BG, Halkier BA, Kliebenstein DJ (2007) Identifying the molecular basis of QTLs: eQTLs add a new dimension. Trends Plant Sci 13:72–77

Holloway B, Li B (2010) Expression QTLs: applications for crop improvement. Mol Breed 26:381–391

Hovav R, Chaudhary B, Udall JA, Flagel L, Wendel JF (2008a) Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. Genetics 179:1725–1733

Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel LE, Wendel JF (2008b) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proc Natl Acad Sci USA 105:6191–6195

Hovav R, Udall JA, Hovav E, Rapp R, Flagel LE, Wendel JF (2008c) A majority of cotton genes are expressed in single-celled fiber. Planta 227:319–329

Hsieh Y-L (1999) Structural development of cotton fibers and linkages to fiber quality. In: Basra AS (ed) Cotton fibers. Food Products Press, Binghampton, pp 137–166

Hubner N, Yagil C, Yagil Y (2006) Novel integrative approach to the identification of candidate genes in hypertension. Hypertension 47:1–5

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17:388–391

Jordan MC, Somers DJ, Banks TW (2007) Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. Plant Biotechnol J 5:442–453

Kang HM, Ye C, Eskin E (2008) Accurate discovery of expression quantitative trait loci under confounding from spurious and genuine regulatory hotspots. Genetics 180:1909–1925

Keurenjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhill D, Koorneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci USA 104:1708–1713

Kirst M, Basten CJ, Myburg AA, Zeng ZB, Sederoff RR (2005) Genetic architecture of transcript-level variation in differentiating xylem of a Eucalyptus hybrid. Genetics 169:2295–2303

Kirst M, Myburg AA, De Leon JPG, Kirst ME, Scott J, Sederoff RR (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of Eucalyptus. Plant Physiol 135:2368–2378

Kliebenstein DJ (2008) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphims or eQTLs. Ann Rev Plant Biol 60:93–114

Kliebenstein DJ, West MA, van Leeuwen H, Loudet O, Doerge RW, St Clair DA (2006) Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinf 7: 308

Lacape J-M, Jacobs J, Arioli T, Derijcker R, Forestier-Chiron JeanJ, Llewellyn D, Thomas E, Viot C (2009) A new inter-specific, Gossypium hirsutum × G. barbadense, RIL population: towards a unified consensus linkage map of tetraploid cotton. Theor Appl Genet 119:281–292

Lacape J-M, Llewellyn D, Jacobs J, Al-Ghazi Y, Liu S, Arioli T, Palai O, Georges S, Giband M, de Asunção H, Viot C, Jean J, Claverie M, Gawryziak G, Vialle M (2010) Meta-analysis of cotton fiber quality QTL across diverse environments in an inter-specific Gossypium hirsutum × G barbandense RIL population. BMC Plant Biol 10:132

Leng C, Li F, Chen G, Liu C (2007) cDNA-AFLP analysis of somatic embryogenesis at early stage in TM-1 (Gossypium hirsutum L.). Acta Bot Boreal Occid Sin 27:233–237

Lin L, Pierce GJ, Bowers JE, Estill JC, Compton RO, Rainville LK, Kim C, Lemke C, Rong J, Tang H, Wang X, Braidotti M, Chen AH, Chicola K, Collura K, Epps E, Golser W, Grover C, Ingles J, Karunakaran S, Kudrna D, Olive J, Tabassum N, Um E, Wissotski M, Yu Y, Zuccolo A, ur Rahman M, Peterson DG, Wing RA, Wendel JF, Paterson AH (2010) A draft physical map of a D-genome cotton species (Gossypium raimondii). BMC Genomics 11:395

Liu R, Wang B, Guo W, Wang L, Zhang T (2011) Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F(2). Theor Appl Genet 123:439–454

Ma X, Xing C, Guo L, Gong Y, Wang H, Zhao Y, Wu J (2008) Analysis of differentially expressed genes in genic male sterility cotton (Gossypium hirsutum L.) using cDNA-AFLP. J Genet Genomics 34:536–543

May LO (1999) Genetic variation in fiber quality. In: Basra AS (ed) Cotton fibers. Food Products Press, Binghampton, pp 183–229

Michaelson JJ, Loguercio S, Beyer A (2009) Detection and interpretation of expression quantitative trait loci (eQTL). Methods 48:265–276

Paterson AH, Rong J, Gingle AR, Chee P, Dennis ES, Llewellyn D, Dure LS III, Haigler C, Myers GO, Peterson DG, Rahman M, Zafar Y, Reddy U, Saranga Y, Mc Stewart J, Udall JA, Waghmare VN, Wendel JF, Wilkins T, Wright RJ, Zaki E, Hafez EE, Zhu J (2010) Sequencing and utilization of the Gossypium genomes. Trop Plant Biol 3:71–74

Pérez-Enciso M, Quevedo JR, Bahamonde A (2007) Genetical genomics: use all data. BMC Genomics 8:69

Ponsuksili S, Murani E, Phatsara C, Schwerin M, Schellander K, Wimmers K (2010) Expression quantitative trait loci analysis of genes in porcine muscle by quantitative real-time RT-PCR compared to microarray data. Heredity 1:1–9

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008) Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. Plant J 53:90–101

Qin L, Prins P, Helder J (2006) 8. Linking cDNA-AFLP-based gene expression patterns and ESTs. In: Liang P, Meade JD, Pardee AB (eds) Methods in molecular biology, vol 317: differential display methods and protocols. Humana Press Inc., Totowa, pp 123–138

Reijans M, Lascaris R, Groeneger AO, Wittenberg A, Wesselink E, van Oeveren J, de Wit E, Boorsma A, Voetdijk B, van der Spek H, Grivell LA, Simons G (2003) Quantitative comparison of cDNA-AFLP, microarrays, and GeneChip expression data in Saccharomyces cerevisiae. Genomics 82:606–618

Rombauts S, Van de Peer Y, Rouzé P (2003) AFLPinSilico, simulating AFLP fingerprints. Bioinformatics 19:776–777

Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee P, Draye X, Saranga Y, Wright RJ, Wilkins TA, May LO, Smith CW, Gannaway JR, Wendel JF, Paterson AH (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577–2588

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Schadt EE, Monks SA, Drake DA, Lusis AL, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and human. Nature 422:297–302

Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Luebberstedt T (2007) Identification of candidate genes

associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population. BMC Genomics 8:22

Torres TT, Metta M, Ottenwälder B, Schlötterer C (2010) Gene expression profiling by massively parallel sequencing. Genome Res 18:172–177

Vuylsteke M, Daele H, Vercauteren A, Zabeau M, Kuiper M (2006) Genetic dissection of transcriptional regulation by cDNA-AFLP. Plant J 45:439–446

Vuylsteke M, Peleman J, van Eijk MJT (2007) AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protoc 2:1399–1413

Wan CY, Wilkins TA (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (Gossypium hirsutum L.). Anal Biochem 223:7–12

Wang J, Yu H, Xie W, Xing Y, Yu S, Xu C, Li X, Xiao J, Zhang Q (2010) A global analysis of QTLs for expression variations in rice shoots at the early seedling stage. Plant J 63:1063–1074

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:140–186

West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript level variation in Arabidopsis. Genetics 175:1441–1450

Xu Z, Kohel RJ, Song GL, Cho JM, Alabady M, Yu J, Koo P, Chu J, Yu SX, Wilkins TA, Zhu YX, Yu JZ (2008) Gene-rich islands for fiber development in the cotton genome. Genomics 92:173–183

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20